Using ADOPT Algorithm and Operational Data to Discover Precursors to Aviation Adverse Events

Vijay Janakiraman, Bryan Matthews and Nikunj Oza

Data Sciences Group

NASA Ames Research Center, CA, USA

AIAA Science and Technology Forum and Exposition (SciTech 2018)

Session: IS-15, Learning, Reasoning, and Data-Driven Systems I

11th January 2018

Outline

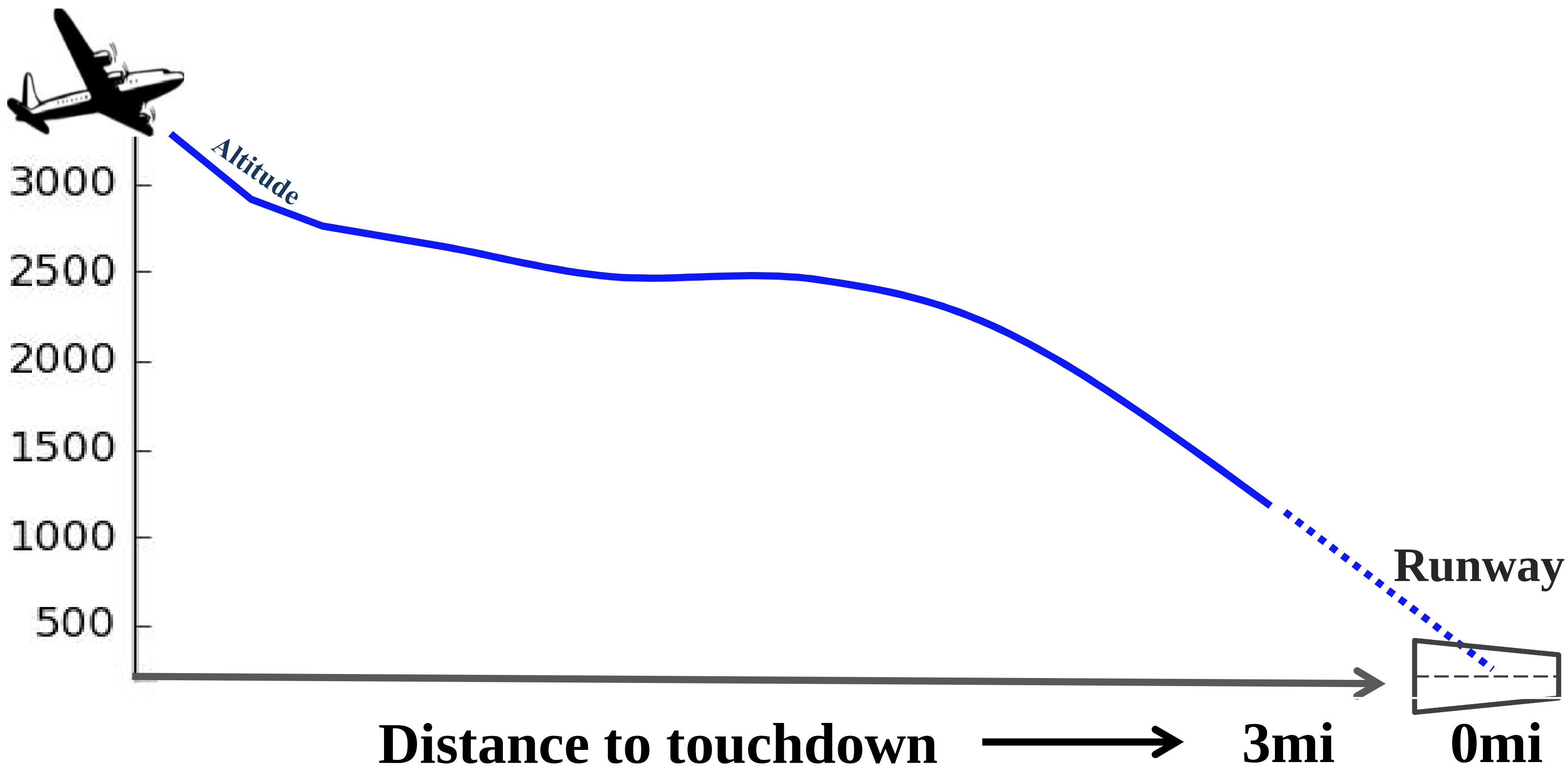
- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

Outline

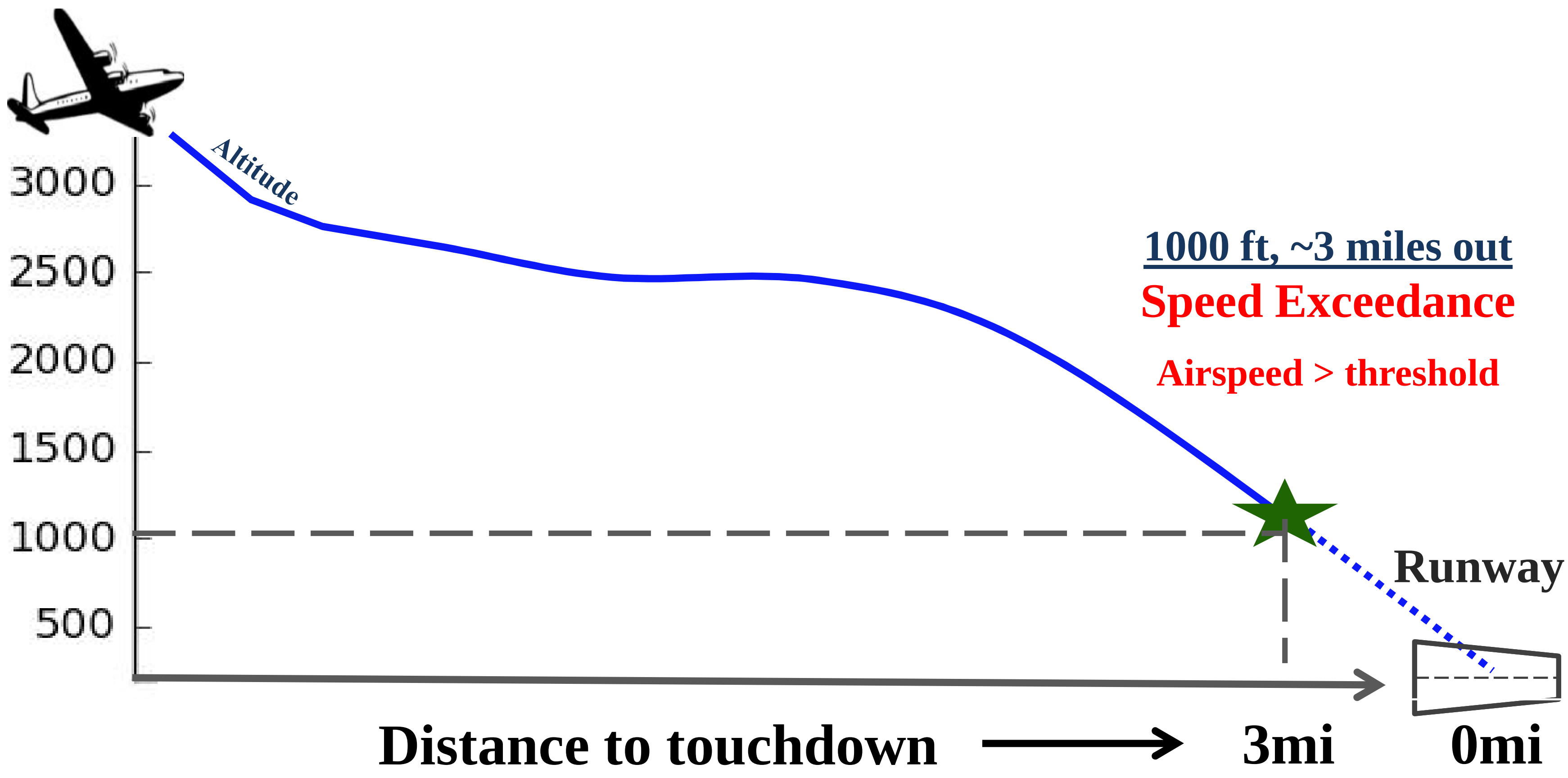
- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

Precursor discovery

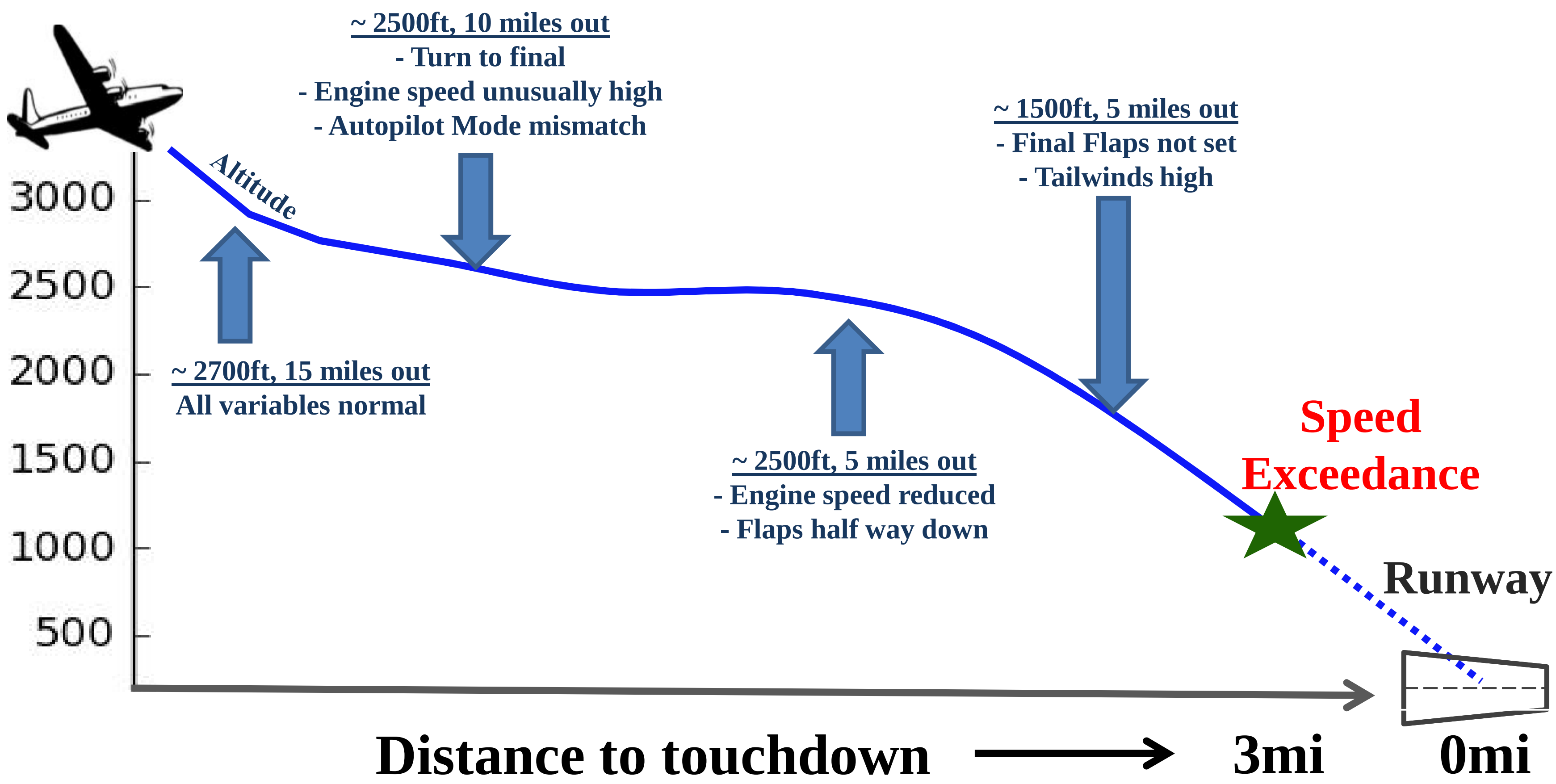
Precursor discovery



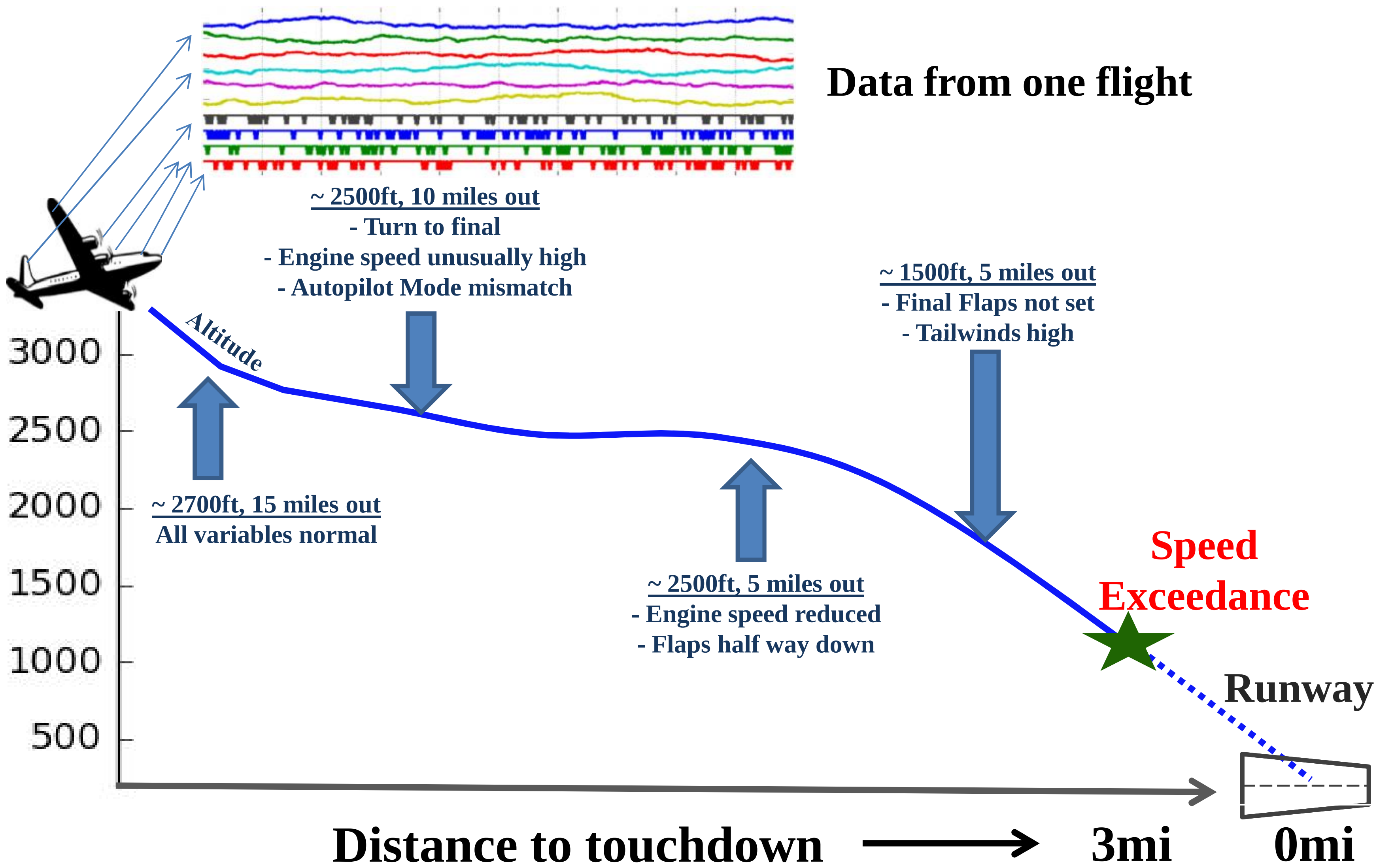
Precursor discovery



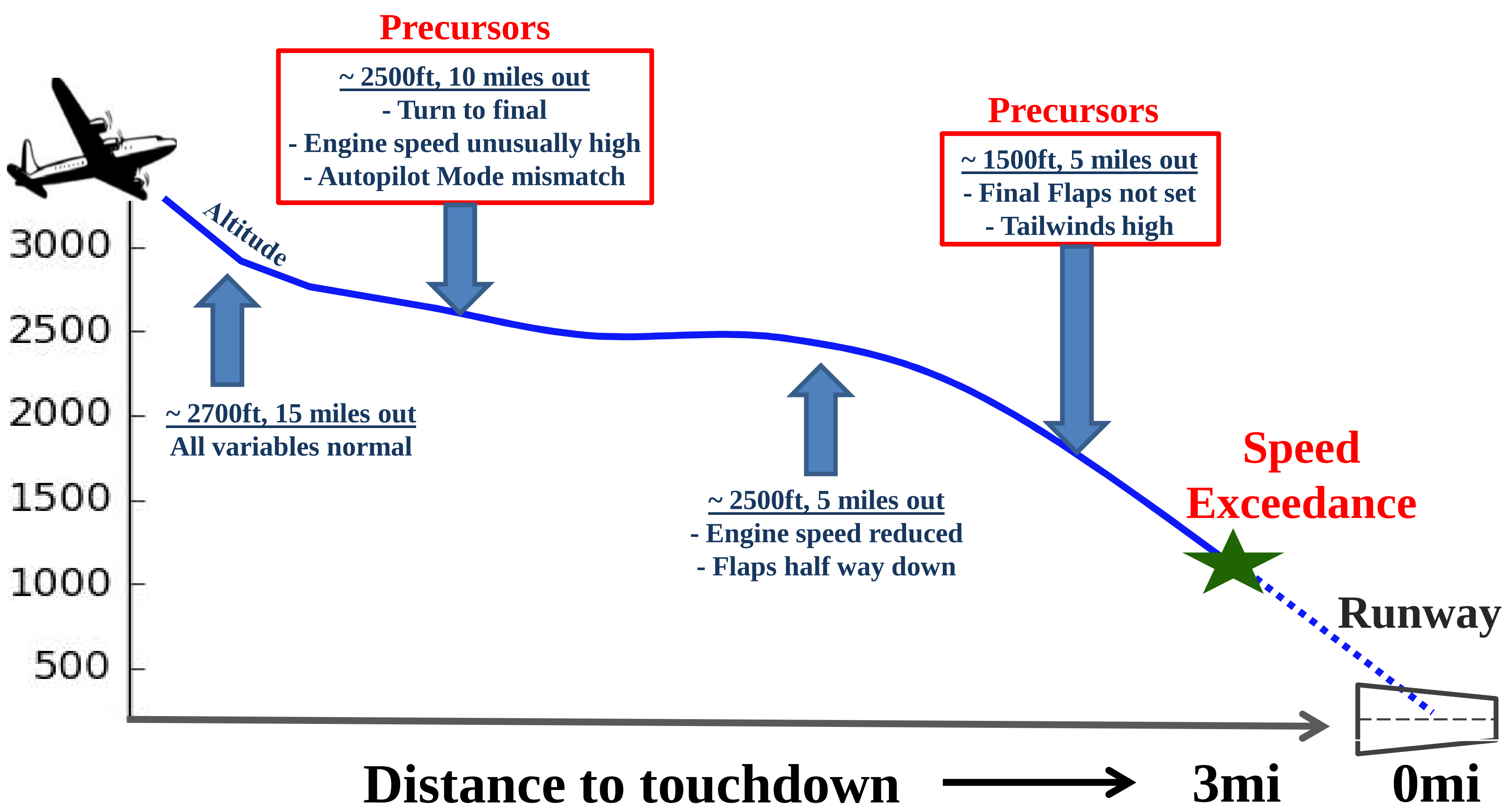
Precursor discovery



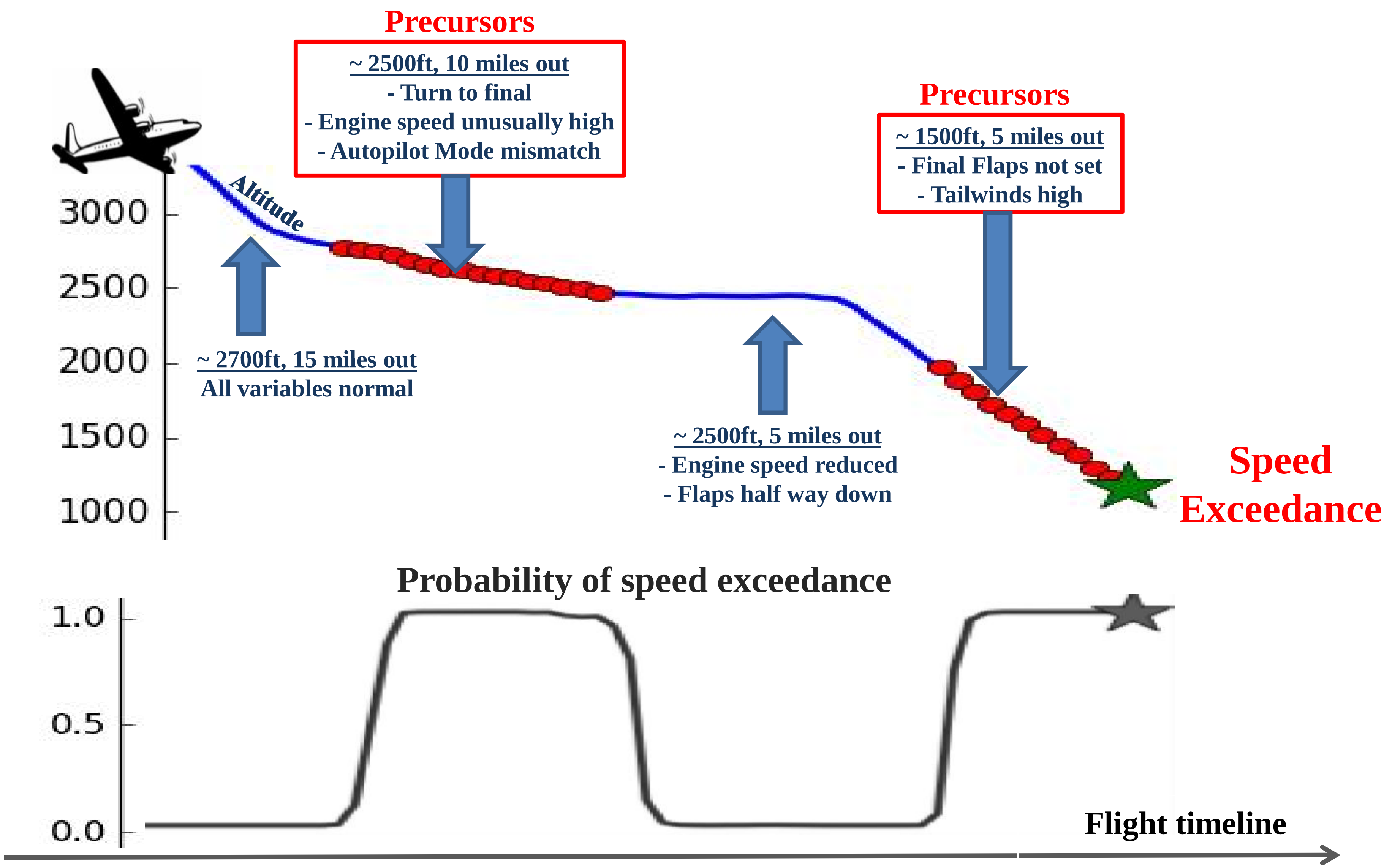
Precursor discovery



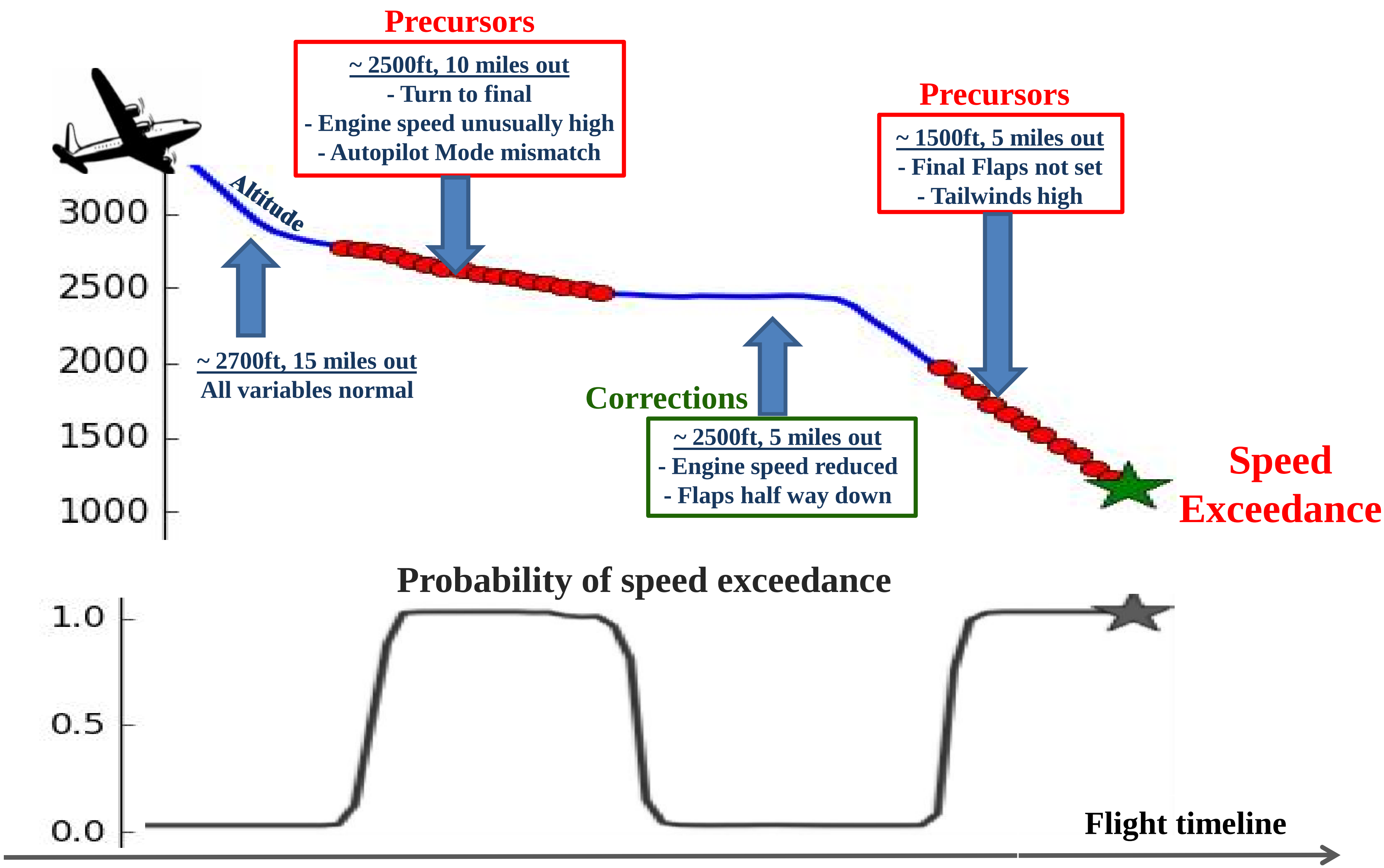
Precursor discovery



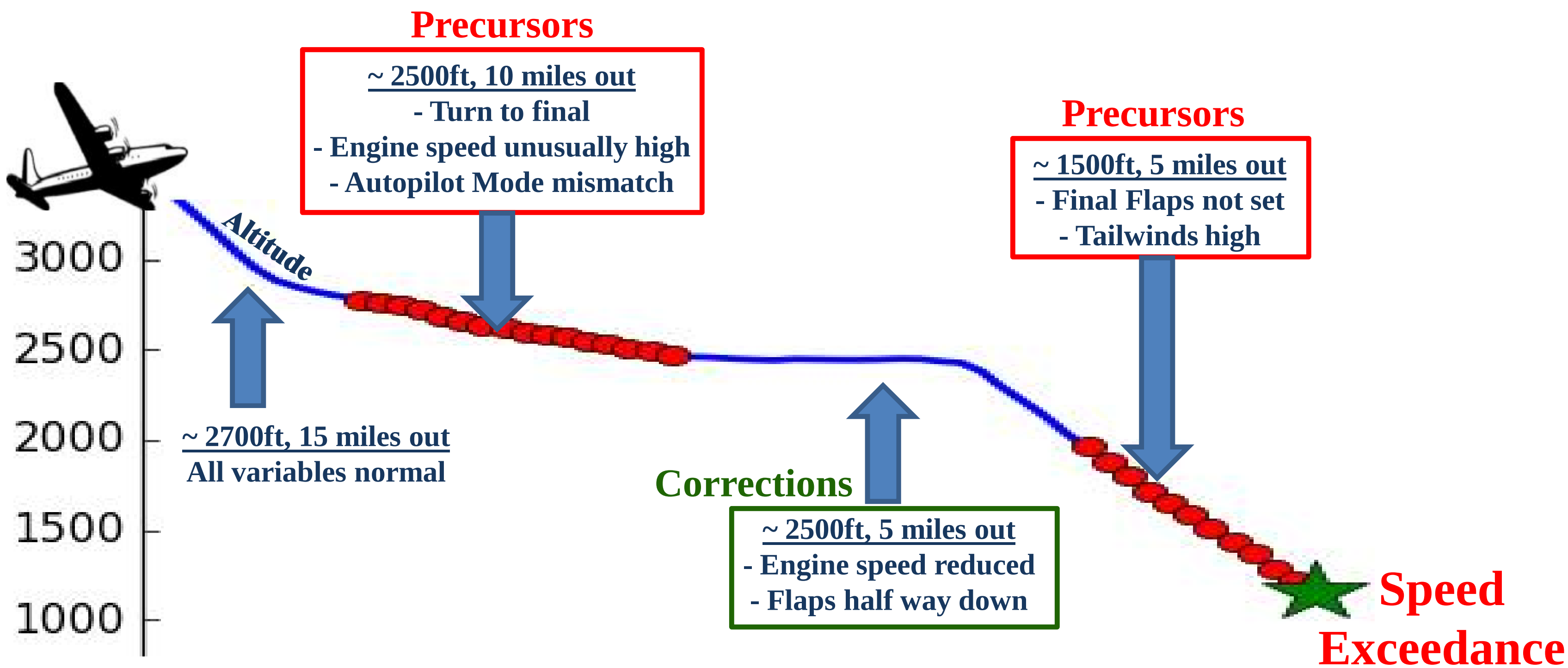
Precursor discovery



Precursor discovery



Precursor discovery



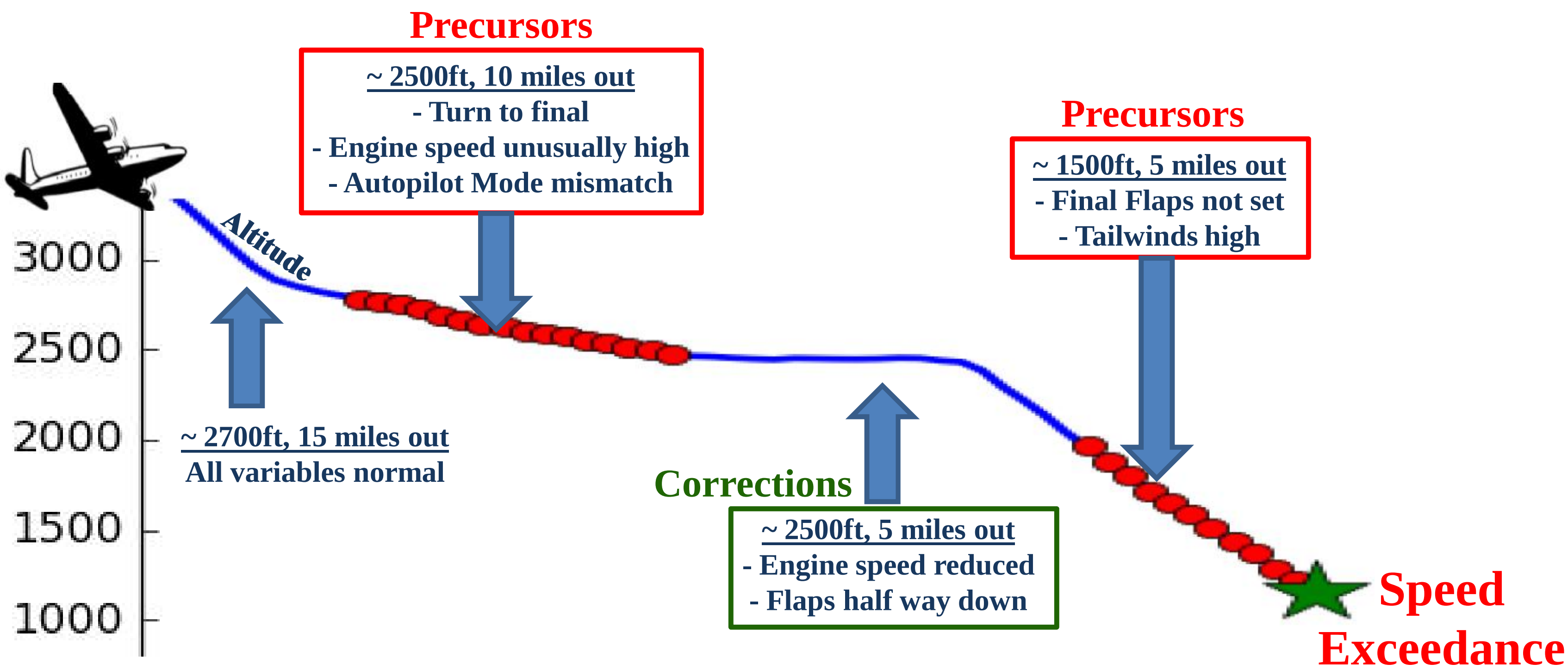
Adverse event may be any event of interest

- Single flight safety events such as exceedances, go-around, stall,
- Multi-flight safety events such as loss of separation, TCAS events,
- Airspace or NAS level events such as GDP, congested sectors, delays,
- Performance events such as high throughput, mission success,

Why find precursors?

Forensic analysis of past events

- Accident investigation
- Hazard identification
- Operations



Why find precursors?

Real-time decision support

- Crew alerting, Situational awareness, Action recommendation

Why find precursors?

- **Forecasting adverse events better and earlier**
  - Generate a knowledge base (precursors)
- **Develop decision support tools**
  - Alerting systems
  - Recommendation systems on corrective actions
- **Improve operator training**
  - Response and recovery from precursors
- **Predictive maintenance**
  - Precursors to component failures

Challenges in Precursor Discovery

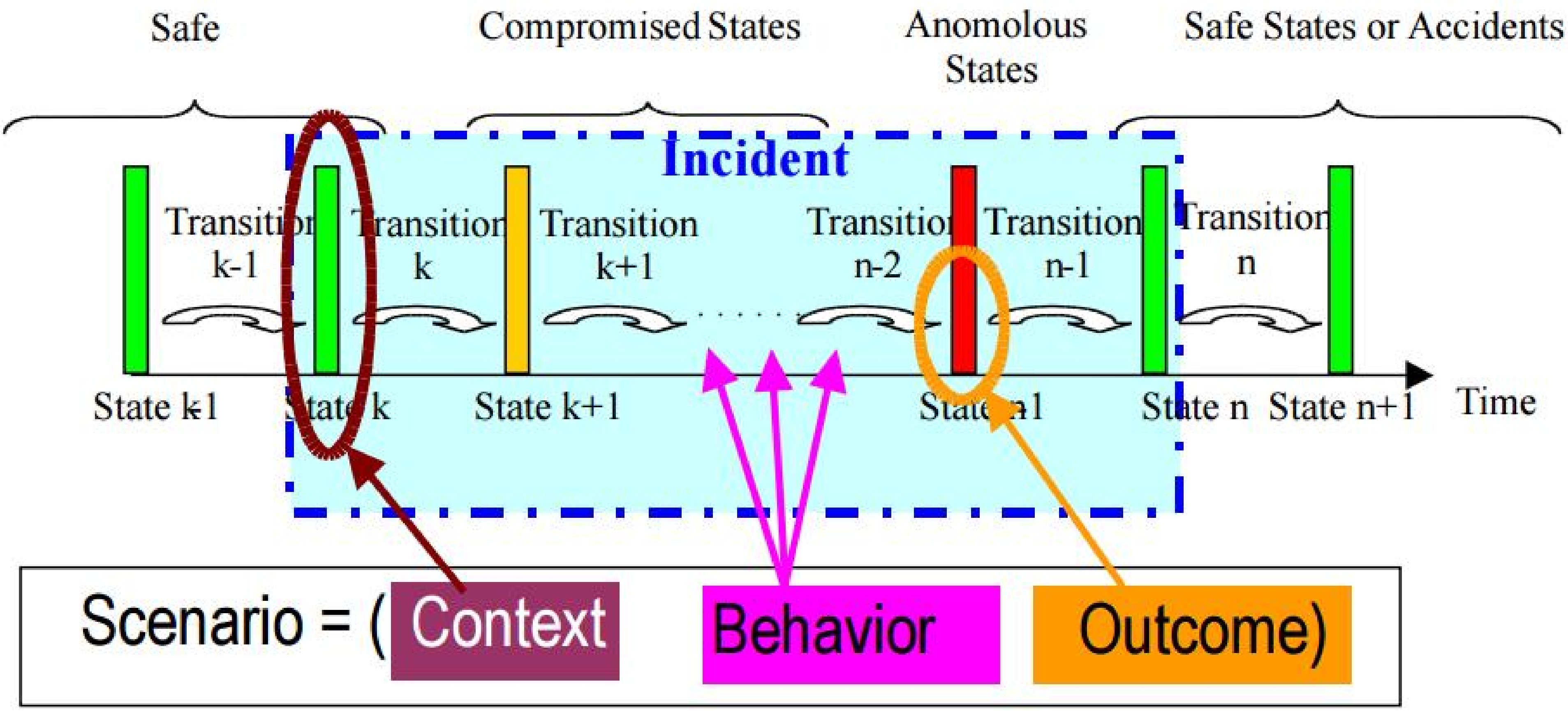
- **Human expert analysis is not scalable**
  - Not easy to find patterns in 100s of time series.
  - Visualization is almost impossible.
  - Subjective variations among experts
  - Costly and slow

- **Data mining is not easy**
  - High dimensions (100s of variables)
  - High velocity of data (1000s of flights per day)
  - Data heterogeneity (continuous, categorical, text, voice, video)
  - Precursors are unlabeled.

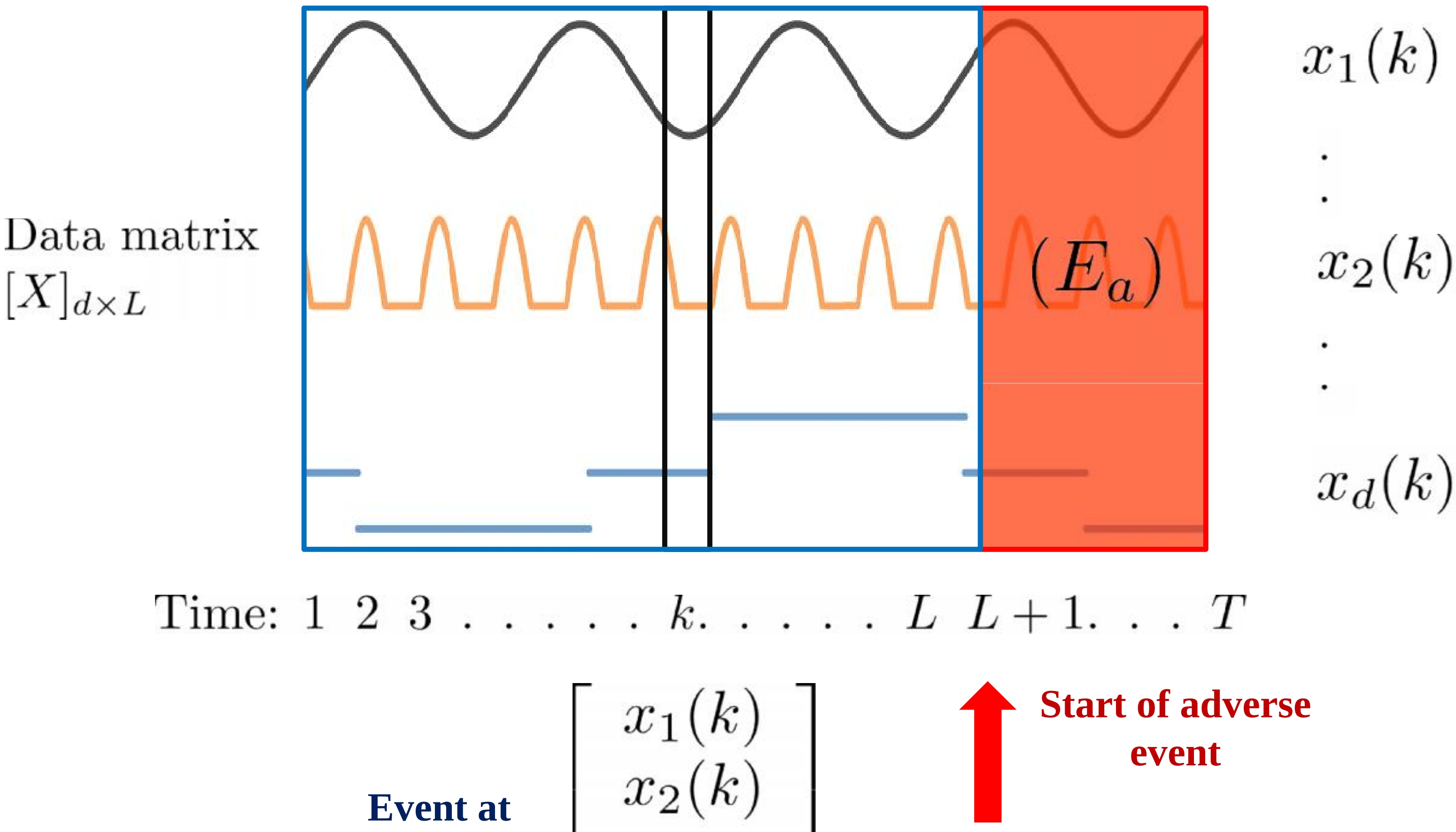
Outline

- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

Anatomy of a Safety Event



Precursor discovery using data



$$\begin{bmatrix} x_1(k) \\ x_2(k) \\ . \\ x_d(k) \end{bmatrix}$$

Problem setup

- Data
  - Adverse time series data $\overline{\mathcal{N}} = \{X_i\},\ i = 1, 2, ..\overline{N};$
  - Nominal time series data $\mathcal{N} = \{X_i\},\ i = 1, 2, ..N;$
  - Unsupervised
- Event
  - A time slice of data $\begin{bmatrix} x_1(k) \\ x_2(k) \\ . \\ x_d(k) \end{bmatrix}$
  - Data is a sequence of events $X_i = [\mathbf{x(1)}, \mathbf{x(2)}, .., \mathbf{x(L_i)}]_i$

Precursor Definition

Given a sequence of events $X = [\mathbf{x(1)}, \mathbf{x(2)}, .., \mathbf{x(L)}]$, an action is any state transition $a_k : \mathbf{x(k)} \longrightarrow \mathbf{x(k+1)}$ where $1 \leq k \leq L$, then a_k is a precursor to E_A if

$$V(a_k) - V(a_k^*) > \delta.$$

where $\delta > 0$,
a_k^* is the expert's action at k,
$V(z)$ is the value function $\propto P(E_A|z)$

Related work

- Precursor discovery in multivariate time series is a new problem
  - No direct algorithm exists

Challenges

1. Unsupervised (no ground truth on precursors)
2. Temporal (long sequences make it hard)
3. High dimensionality

Possible approaches

<u>Rule Mining</u>
Temporal rule mining [1,2] ① ②
Motif mining [7] ① ②
Clustering [10] ③

<u>Model Based</u>
HMM [9] ① ②
Utility based rules [8] ① ②

<u>Causality</u>
Causal Bayesian models,
Granger causality
[3,4,5,6] ① ②

Issues/Drawbacks

① Computationally expensive (scales combinatorial/exponential with number of items).
② Doesn’t handle continuous data (or needs discretization which grows combinatorial).
③ Similarity metric not easy to define for high dimensional data.

References

- [1] R. Agrawal and R. Srikant, “Mining sequential patterns," in Proceedings of the Eleventh International Conference on Data Engineering, ser. ICDE '95. Washington, DC, USA: IEEE Computer Society, 1995, pp. 3-14.
- [2] G. Das, K. ip Lin, H. Mannila, G. Renganathan, and P. Smyth, “Rule discovery from time series," in KDD 98. AAAI Press, 1998, pp. 16-22.
- [3] K. Karimi and H. Hamilton, “Discovering tempo ral/causal rules: A comparison of methods," in Advances in Artificial Intelligence, ser. Lecture Notes in Computer Science, Y. Xiang and B. Chaib-draa, Eds. Springer Berlin Heidelberg, 2003, vol. 2671, pp. 175-189.
- [4] S. Kleinberg, “Causal inference with rare events in large-scale time-series data," in Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, ser. IJCAI '13. AAAI Press, 2013, pp. 1444-1450.
- [5] A. Arnold, Y. Liu, and N. Abe, “Temporal causal modeling with graphical granger methods," in Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD'07. New York, NY, USA: ACM, 2007, pp. 66-75.
- [6] J. B. D. Cabrera and R. K. Mehra, “Extracting Precursor Rules from Time Series: A Classical Statistical Viewpoint," in SIAM International Conference on Data Mining, 2002.
- [7] A. McGovern, D. Rosendahl, R. Brown, and K. Droegemeier, “Identifying predictive multidimensional time series motifs: an application to severe weather prediction," Data Mining and Knowledge Discovery, vol. 22, no. 1-2, pp. 232{258, 2011.
- [8] G. Maragatham and M. Lakshmi, "A strategy for mining utility based temporal association rules," Trendz in Information Sciences & Computing (TISC), 2010, Chennai, 2010, pp. 38-41.
- [9] Z. Zhu and G. Deng, "Mining Interest Association Rules in Website Based on Hidden Markov Model," Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on, Dalian, 2008, pp. 1-4.
- [10] Rani, Y. Leela Sandhya, P. Naga Deepthi, and Ch Rama Devi. "Clustering Algorithm for Temporal Data

Background: Markov Model

An MDP is a tuple $(\mathcal{S}, \mathcal{A}, P_{s,a}, \gamma, R)$

- $\mathcal{S} = \mathbb{R}^d$ is a continuous state space with d state variables,

- $\mathcal{A} = \mathbb{R}^l$ is an action space with l action variables,

- $\{P^a_{s,s'}\}$ (or $P_{ss'}$ if actions are unknown) are the state transition probabilities,

- $\gamma \in [0, 1]$ is the discount factor,

- $R : \mathcal{S} \to \mathbb{R}$ is the underlying reward function,

$$\text{policy: } \pi(s, a) = p(a|s),$$

$$\text{optimal policy: } \pi_E(s) = a^*,$$

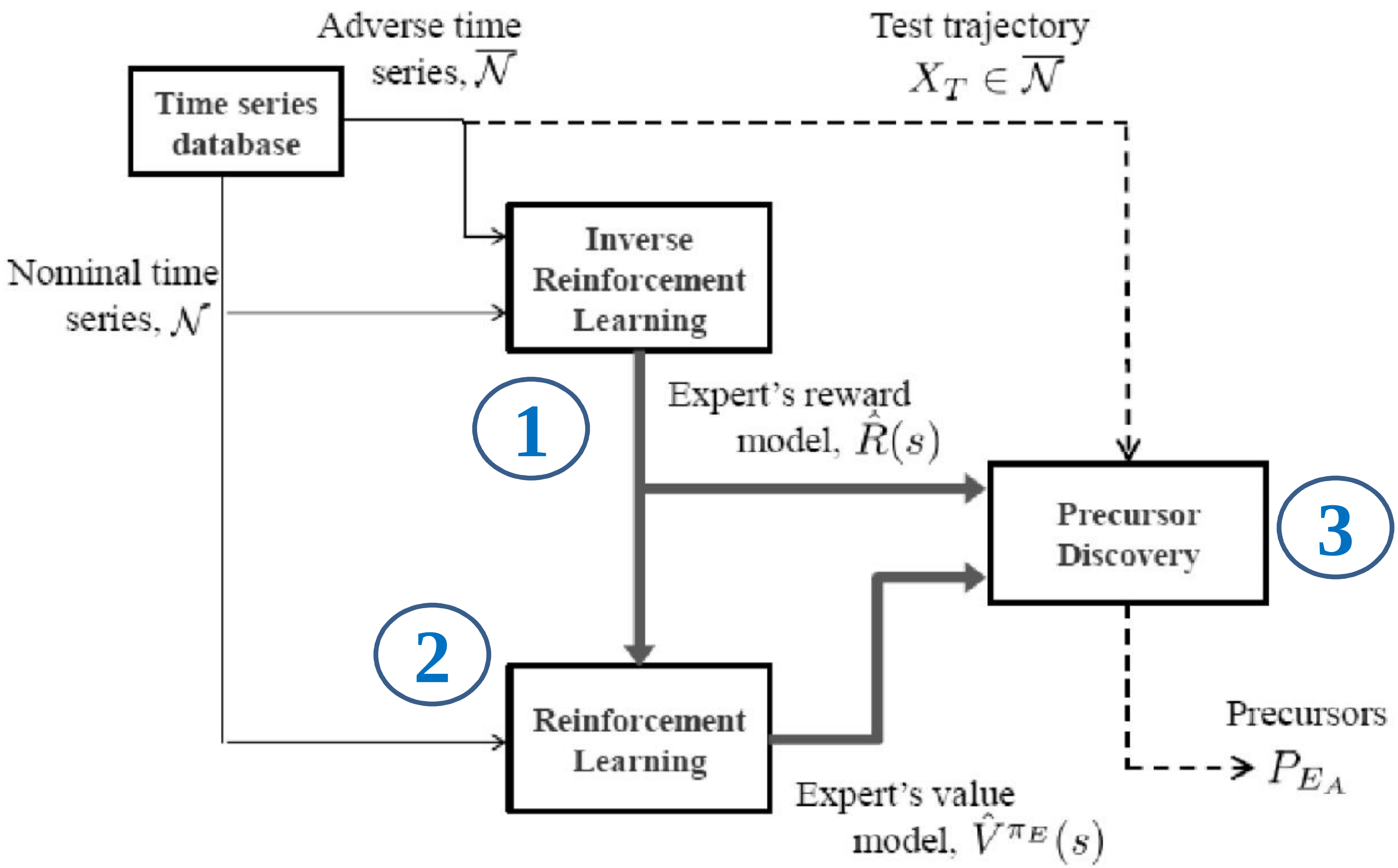
Background: Value Function and Bellman's Optimality

- The value of state $\mathbf{x_0}$ under policy π is
  $V^{\pi}(\mathbf{x_0}) = E[R(\mathbf{x_0}) + \gamma R(\mathbf{x_1}) + .. + \gamma^{L} R(\mathbf{x_L})|\pi]$
  where the expectation is over the distribution of sequences starting from $\mathbf{x_0}$.

- The expert's policy π_E
  $\pi_E(\mathbf{x}) \geq \pi_i(\mathbf{x}) \quad \Longleftrightarrow \quad V^{\pi_E}(\mathbf{x}) \geq V^{\pi_i}(\mathbf{x}) \quad \forall \quad \pi_i$

- Bellman's optimality
  $\pi_E(\mathbf{x_k}) = arg\max_{\{feasible\ \mathbf{x_{k+1}}\}} V^{\pi_E}(\mathbf{x_{k+1}})$

ADOPT Framework



Step 1: Expert's Reward Model

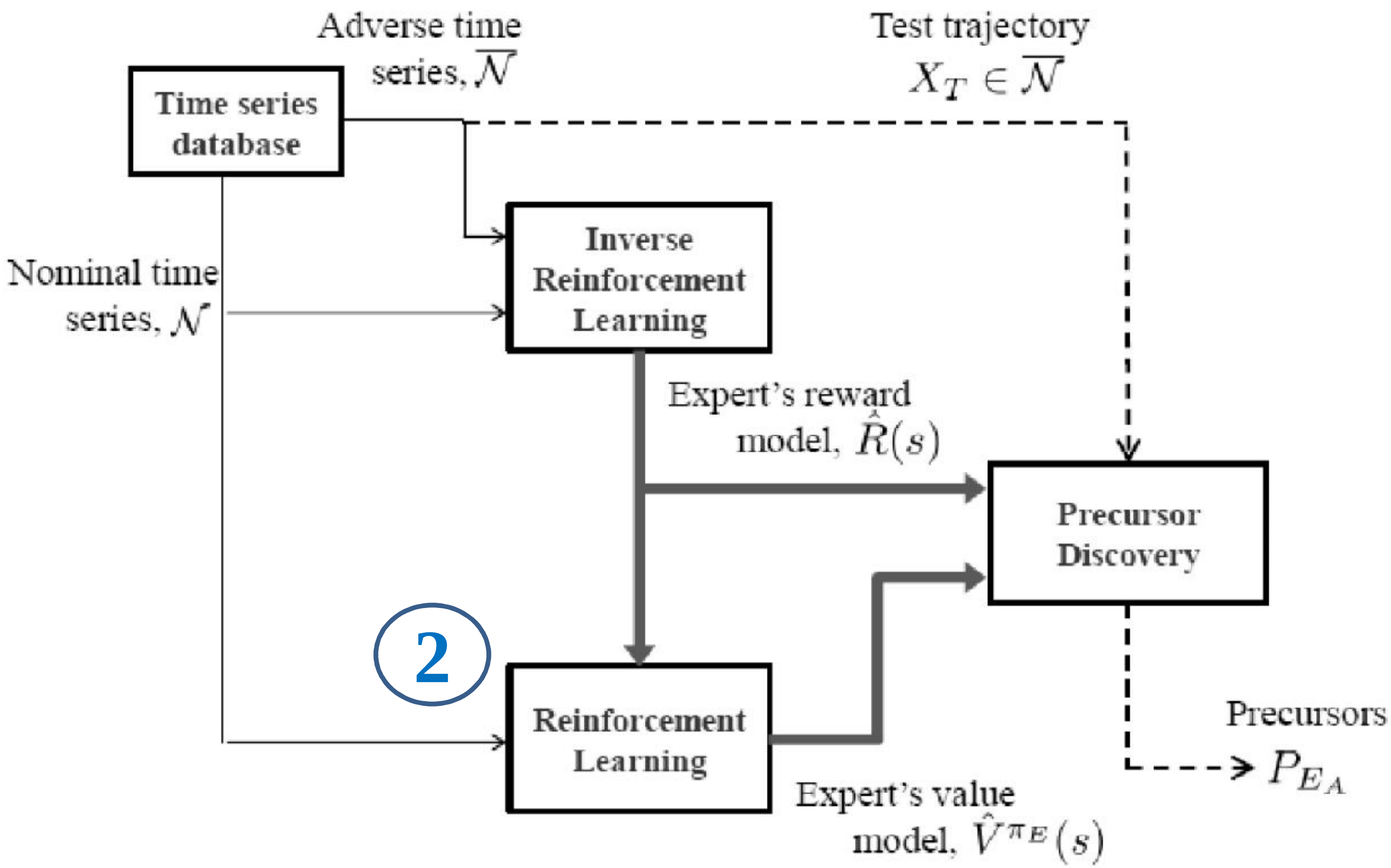
- $R(\mathbf{x};\boldsymbol{\alpha}) = \alpha_1\phi_1(\mathbf{x}) + \alpha_2\phi_2(\mathbf{x}) + .. + \alpha_m\phi_m(\mathbf{x})$
  - A general model of the expert's reward
  - $\boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_m]^T$ to be estimated
  - $\phi_i(\mathbf{x}); i = 1, 2, .., m$ are some known basis functions (gaussian)

- $\boldsymbol{\alpha}^* = arg\min_{\boldsymbol{\alpha}}\{E_{\mathbf{x}_0}[V^{\pi_{adv}}(\mathbf{x}_0;\boldsymbol{\alpha})] - E_{\mathbf{x}_0}[V^{\pi_E}(\mathbf{x}_0;\boldsymbol{\alpha})]\}$
  - such that $|\alpha_i| \leq 1, i = 1, 1, .., m$

A. Y. Ng and S. Russell, "Algorithms for inverse reinforcement learning," ICML 2000.

$$\boxed{\hat{R}(\mathbf{x}) = f_R(\mathbf{x};\boldsymbol{\alpha}^*) = \sum_{i=1}^{m} \alpha_i^* \phi_i(\mathbf{x})}$$

ADOPT Framework

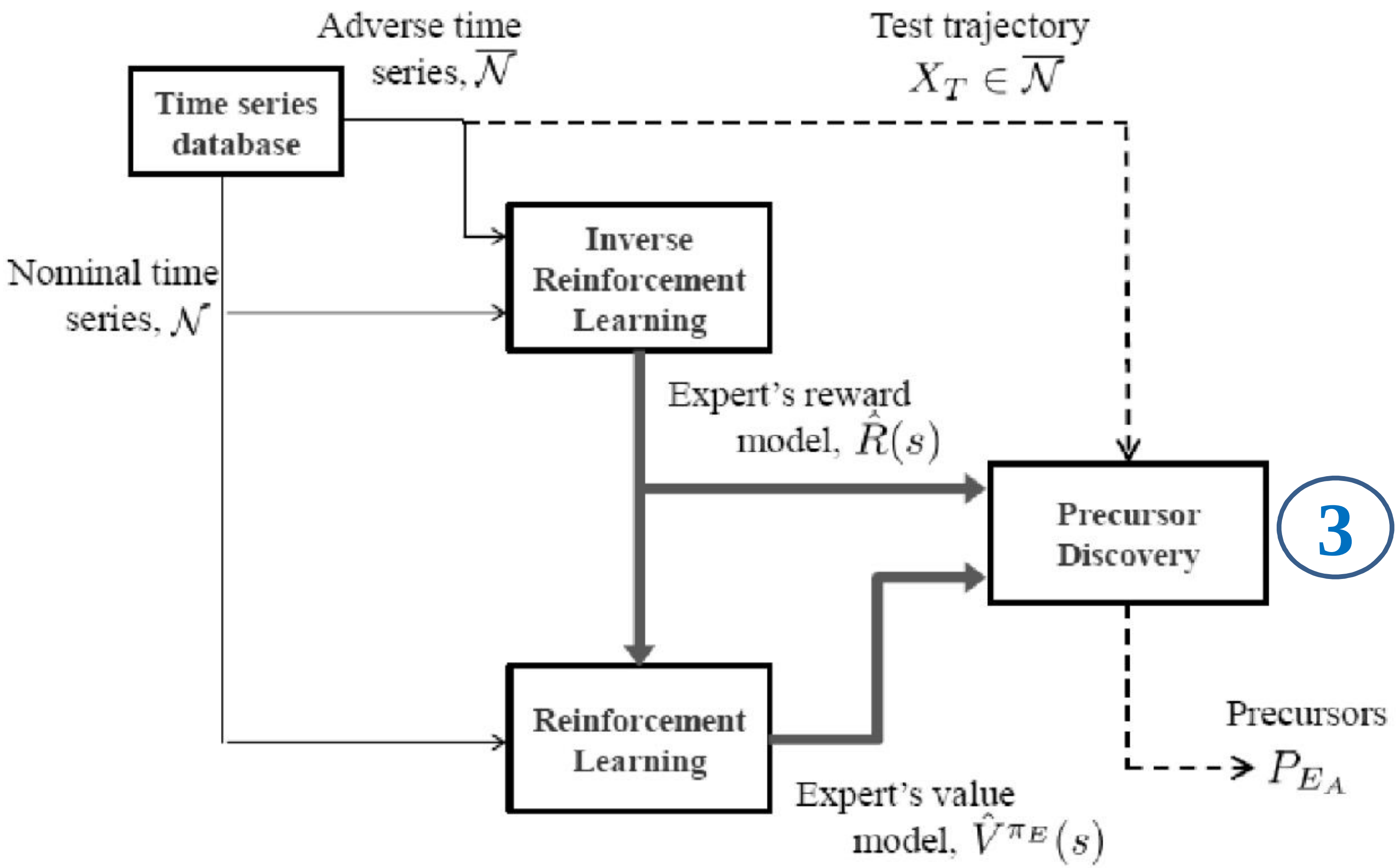


Step 2: Expert's Value Model

- Value estimation by Monte Carlo method
  - Known reward (from previous step)
  - Time series data as Monte Carlo samples
  - Return $Ret(\mathbf{x}) = \sum \gamma^k R(\mathbf{x_k})$ for each state $\mathbf{x}$ as accumulated rewards
- For every labeled pair $(\mathbf{x}_i, Ret(\mathbf{x}_i))$, a regression model $\hat{V}^{\pi_E}(\mathbf{x};\theta)$ parameterized by θ can be built
  - $\theta^* = arg\min_\theta \frac{1}{N_s}\sum_{i=1}^{N_s} \|Ret(\mathbf{x}_i) - \hat{V}^{\pi_E}(\mathbf{x}_i;\theta)\|^2 + \frac{\mu}{2}\|\theta\|^2$

$$\boxed{\hat{V}^{\pi_E}(\mathbf{x}) = f_V(\mathbf{x};\theta^*)}$$

ADOPT Framework



Step 3: Precursor Discovery

Given a sequence of events $X = [\mathbf{x(1)}, \mathbf{x(2)}, .., \mathbf{x(L)}]$, an action is any state transition $a_k : \mathbf{x(k)} \longrightarrow \mathbf{x(k+1)}$ where $1 \leq k \leq L$, then a_k is a precursor to E_A if

Test action

$$V(a_k^*) - V(a_k) > \delta.$$

- requires finding the "optimal" decision
  - Bellman's optimality
    $\mathbf{x^*_{k+1}} = arg\max_{\{feasible\ \mathbf{x_{k+1}}\}} V^{\pi_E}(\mathbf{x_{k+1}})$

- requires scoring the suboptimal decisions
  - $PI_k = V^{\pi_E}(\mathbf{x^*_{k+1}}) - V^{\pi_E}(\mathbf{x_{k+1}})$
  - A weighted contribution from reward may be added to tradeoff short term vs long term precursors

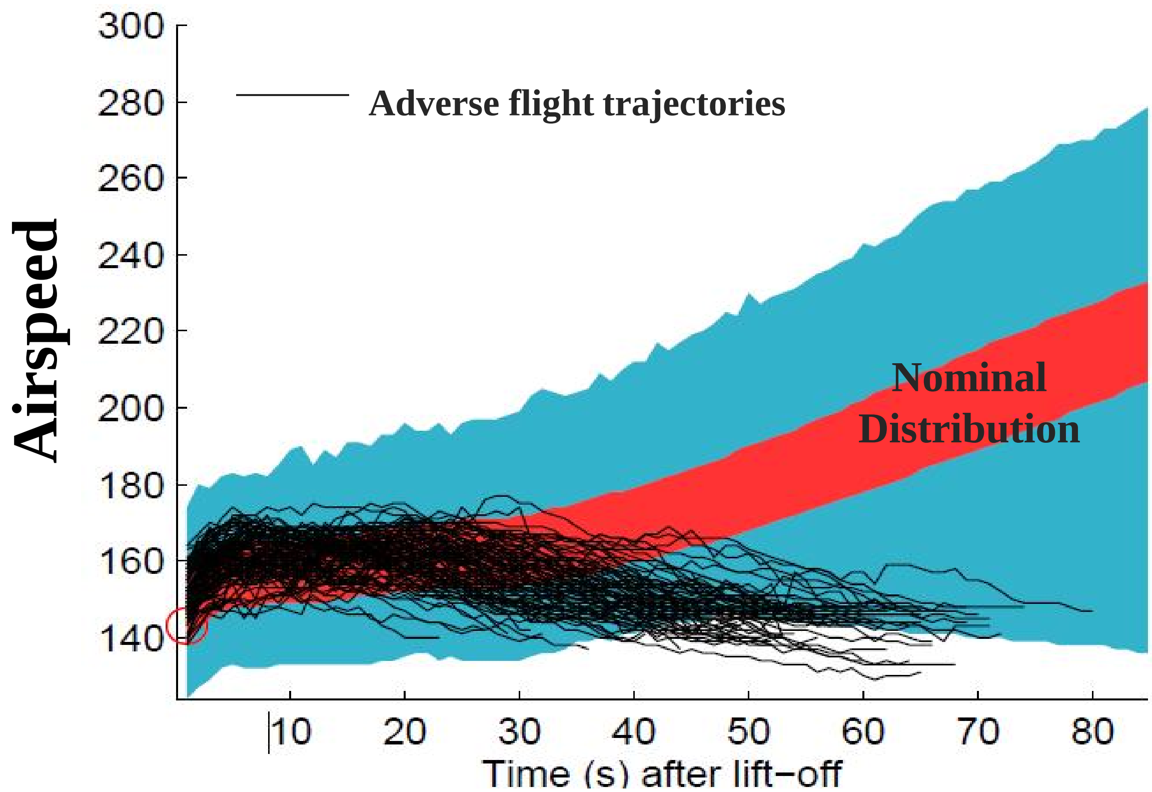
Outline

- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

Take-off Stall Hazard

Adverse event: Drop in airspeed after take-off by at least a 20 knots

Goal: To find precursors using flight recorded data



Outline

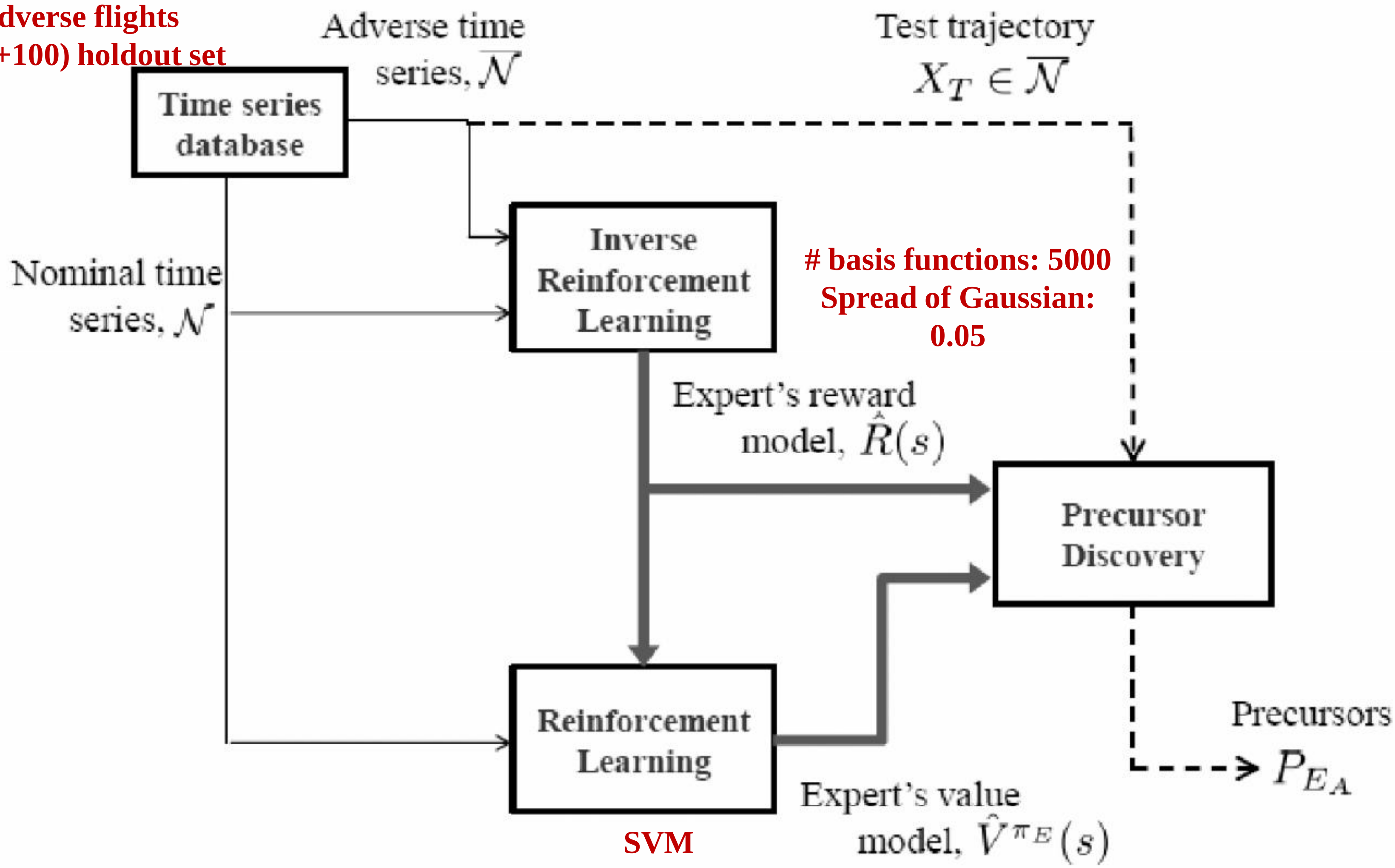
- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

Factors affecting drop in airspeed

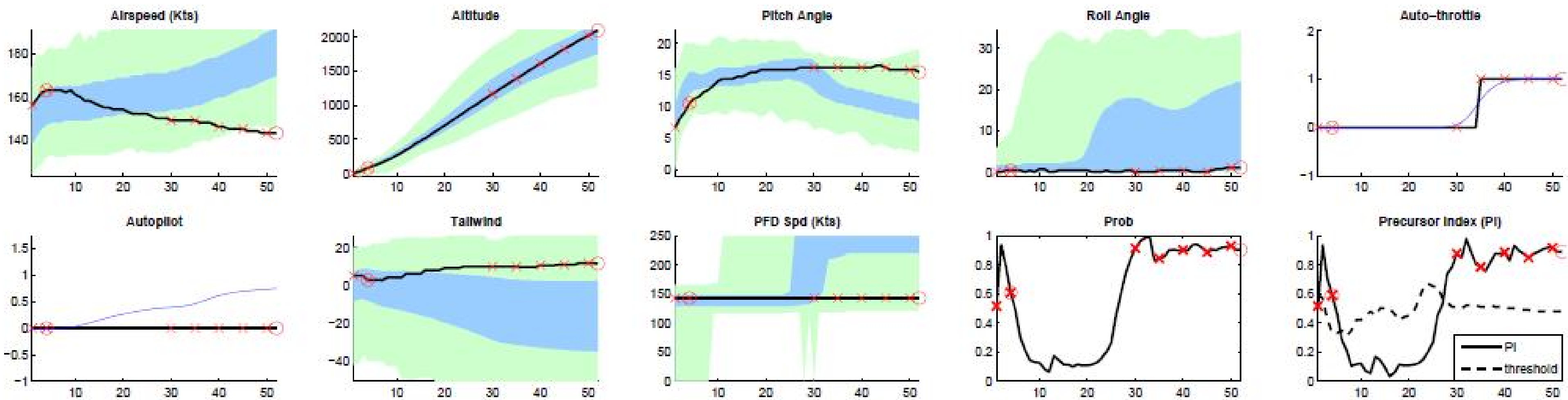
- **Human Factors**
  - Errors in reference speed calculations, estimating AC weight, energy management.
  - human-machine interactions, fatigue, aggressive flying, mode confusion.
- **Environmental**
  - Tail winds, wind shear, sensor failure
- **Procedural**
  - Avoiding terrain, flying over restricted area

ADOPT analysis

400 nominal flights
400 adverse flights
200 (100+100) holdout set

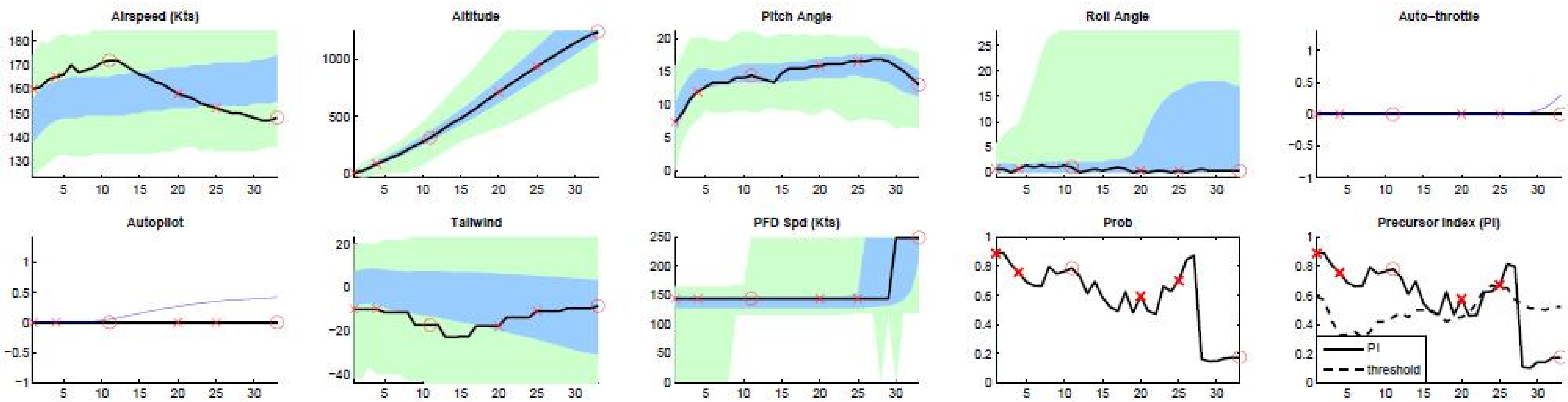


Flight analysis 1 – reference speed set incorrectly



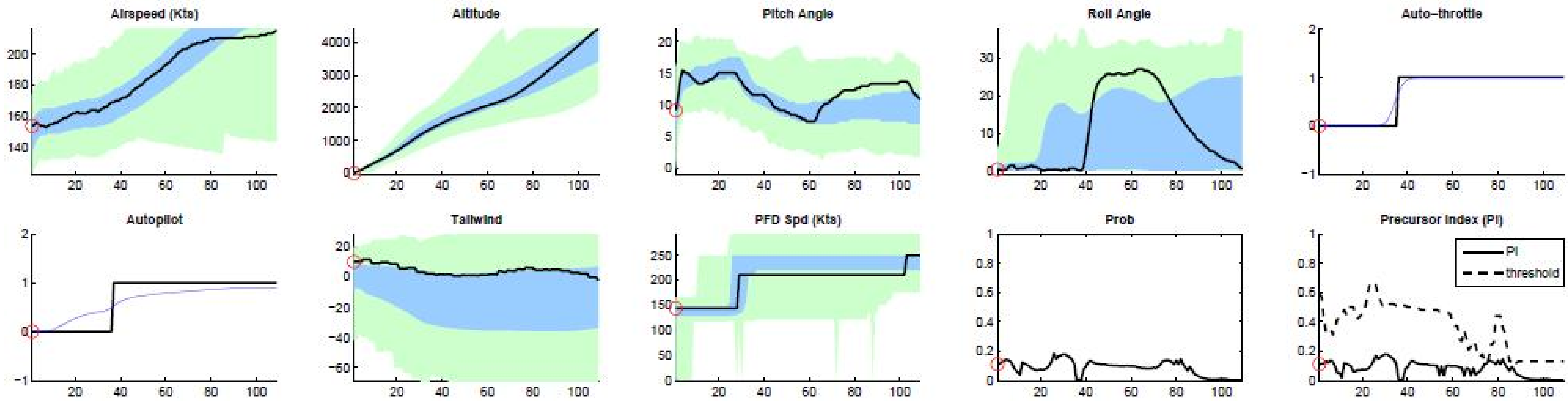
Time = 1s	4s	30s	35s	40s	45s	50s
Tailwind	Pitch Angle	PFD Spd	PFD Spd	Pitch Angle	Pitch Angle	PFD Spd
Pitch Angle	Altitude	Pitch Angle	Tailwind	PFD Spd	PFD Spd	Pitch Angle
Roll Angle	Roll Angle	Auto-throttle	Pitch Angle	Tailwind	Tailwind	Tailwind
Altitude	Tailwind	Roll Angle	Roll Angle	Auto-throttle	Auto-throttle	Auto-throttle
Auto-throttle	Auto-throttle	Altitude	Altitude	Autopilot	Autopilot	Autopilot

Flight analysis 2 – reference speed set incorrectly



Time = 1s	4s	20s	25s
Pitch Angle	Tailwind	Tailwind	Pitch Angle
Tailwind	Pitch Angle	Pitch Angle	PFD Spd
Altitude	Roll Angle	Roll Angle	Tailwind
Roll Angle	Altitude	PFD Spd	Roll Angle
Auto-throttle	Auto-throttle	Altitude	Altitude

Flight analysis 1 – Nominal Flight



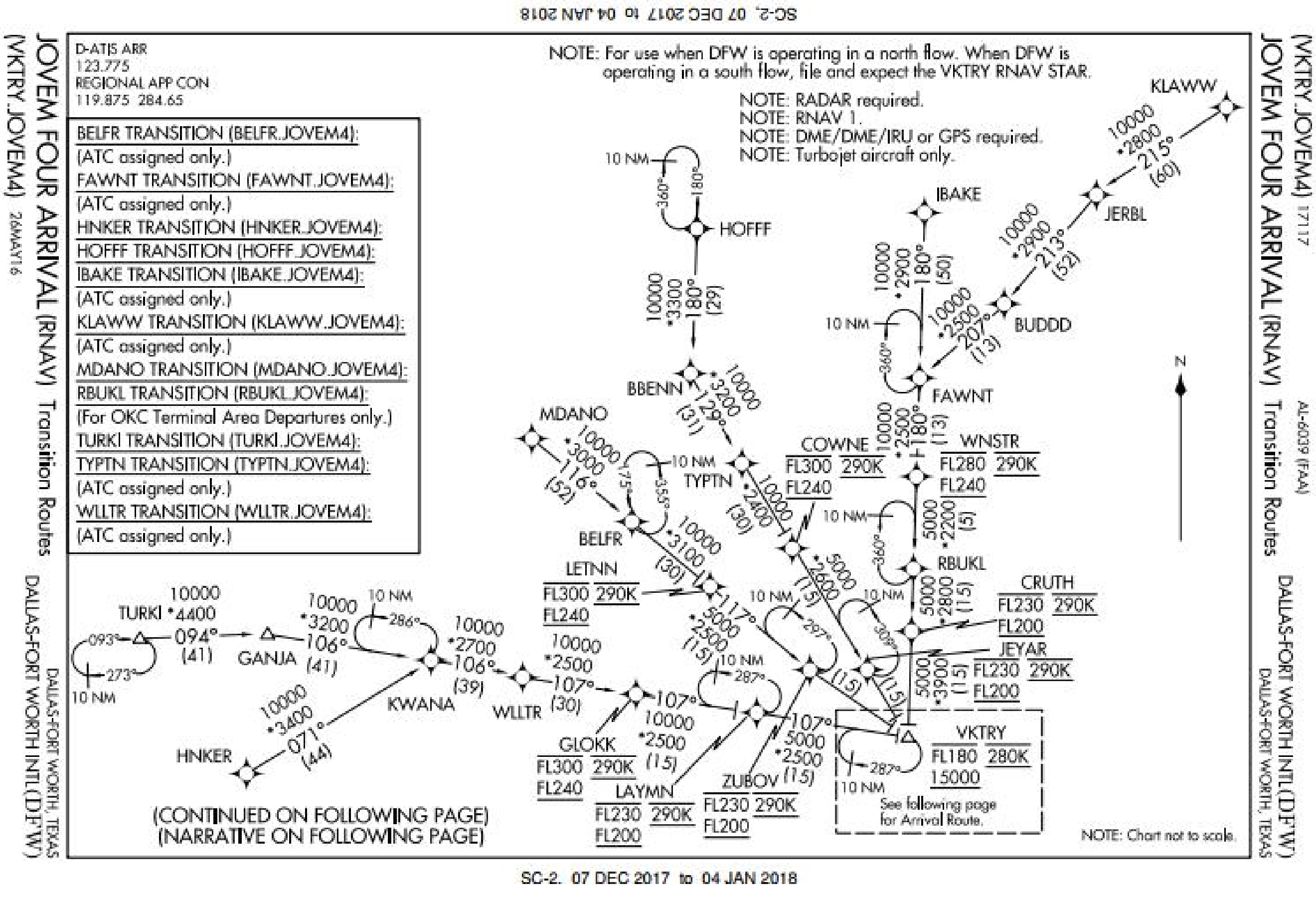
Outline

- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

STAR procedure adherence

Adverse event: Drop in airspeed after take-off by at least a 20 knots

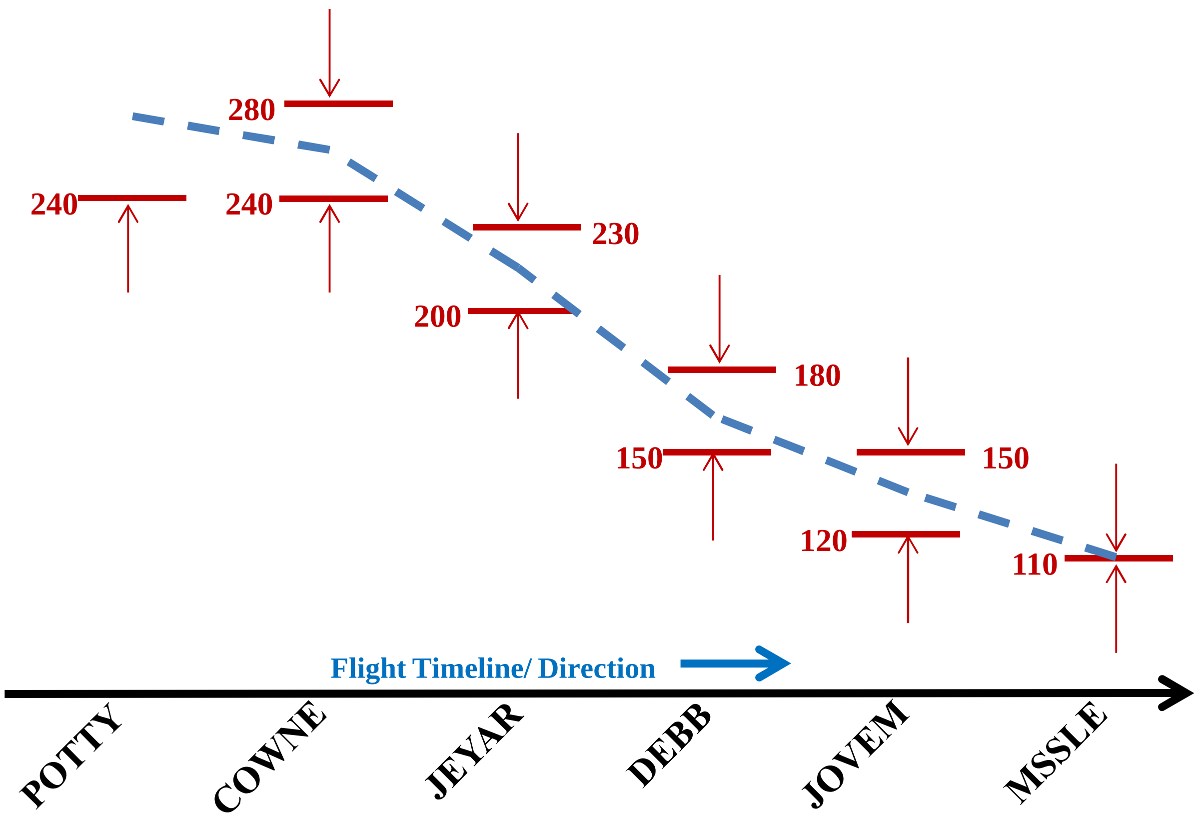
Goal: To find precursors using flight recorded data



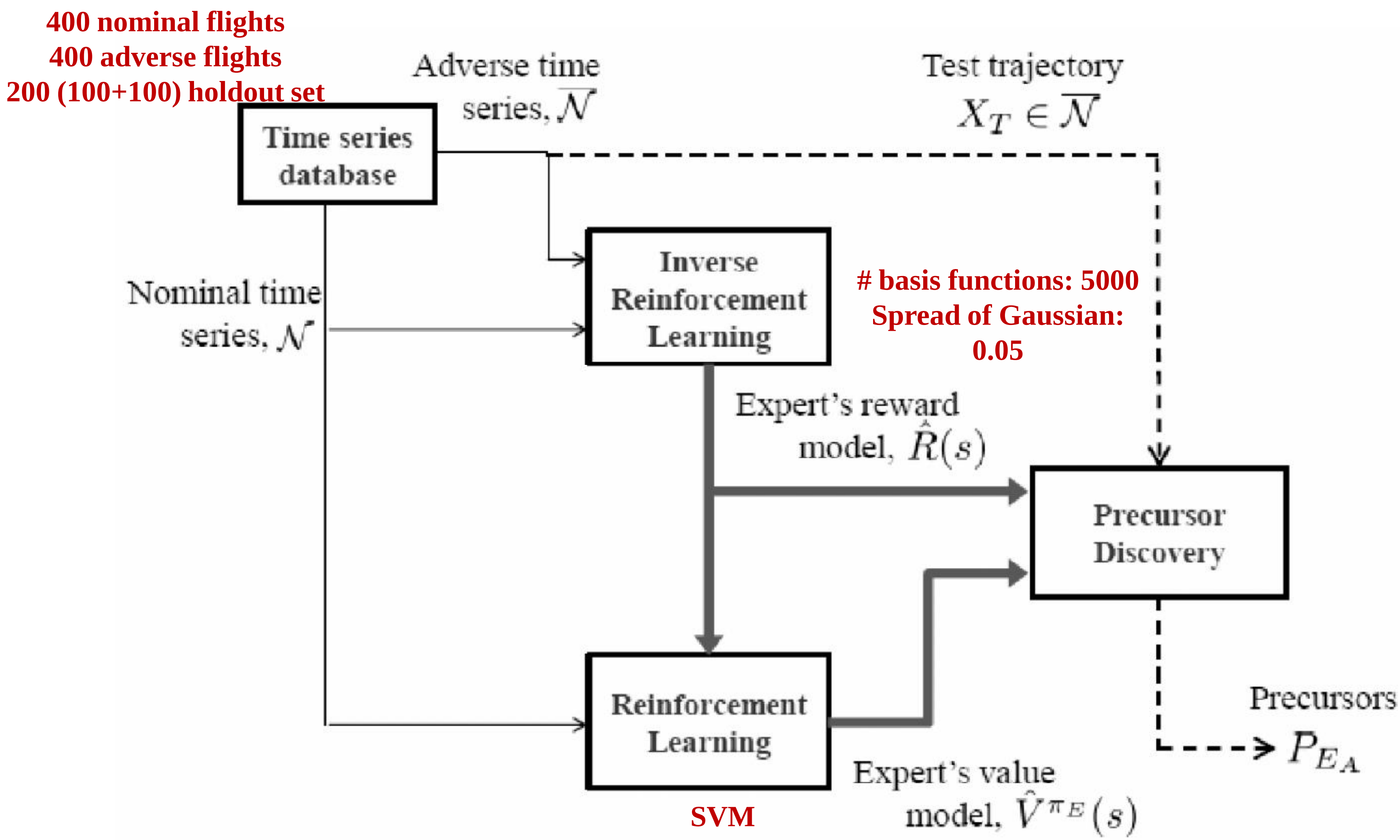
Case Study 2 – STAR procedure adherence

Adverse event: Drop in airspeed after take-off by at least a 20 knots

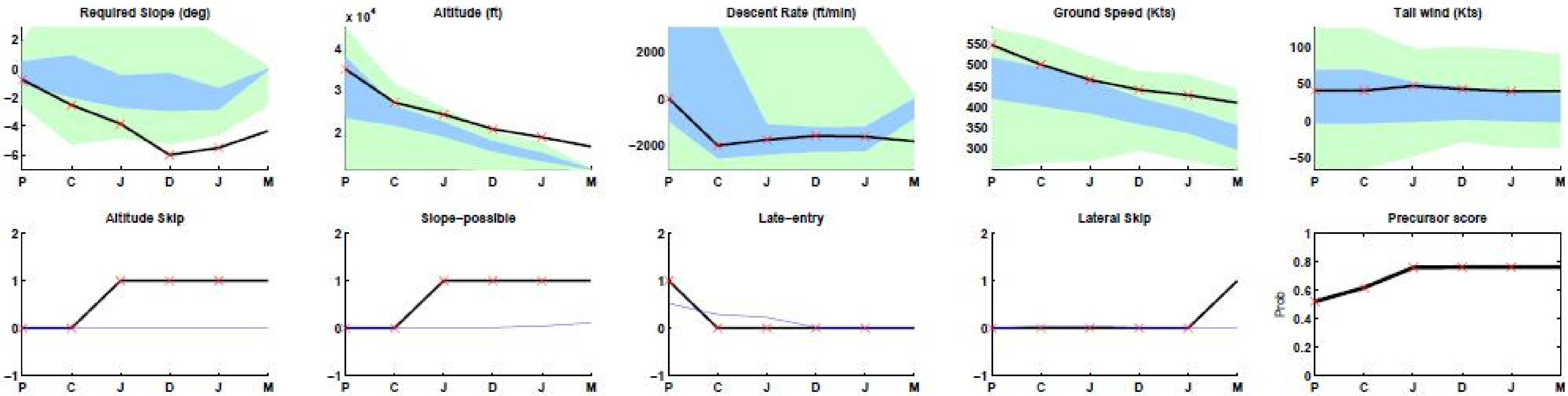
Goal: To find precursors using flight recorded data



ADOPT analysis

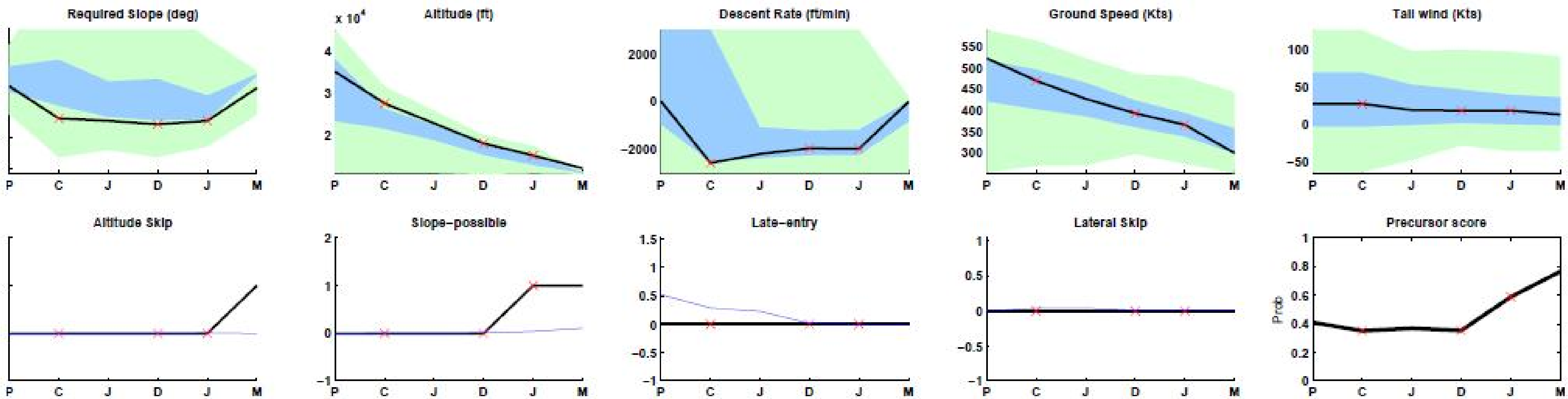


Flight analysis 1 – reference speed set incorrectly



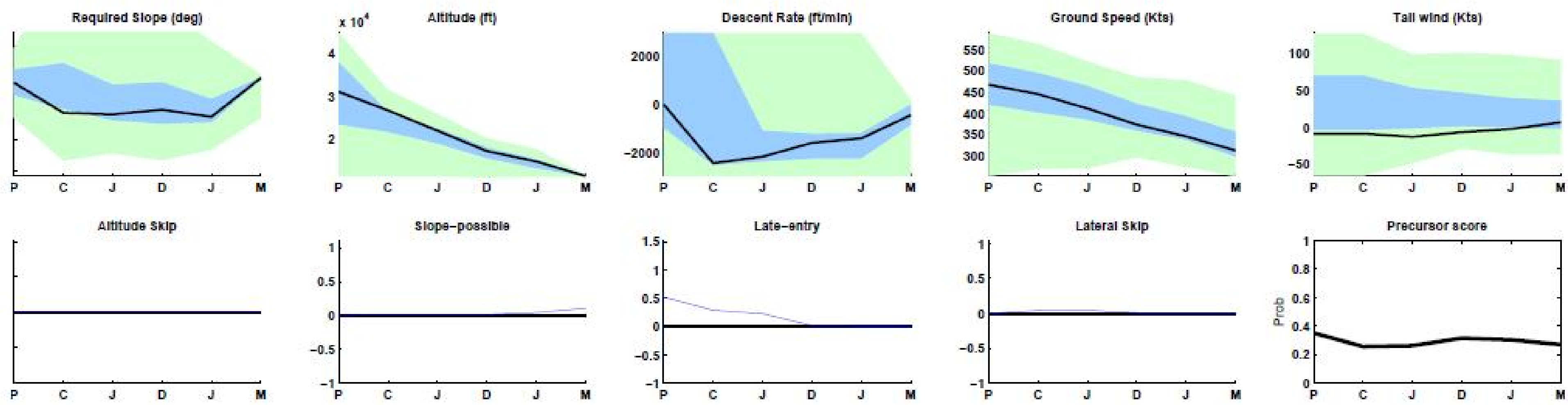
Time = 1s	20s	25s	30s	35s
Tailwind	Pitch Angle	Pitch Angle	Autopilot	Autopilot
Roll Angle	Autopilot	Roll Angle	Roll Angle	Auto-throttle
Pitch Angle	Roll Angle	Tailwind	Auto-throttle	Roll Angle
Altitude	PFD Spd	PFD Spd	Pitch Angle	Tailwind
Auto-throttle	Tailwind	Auto-throttle	Tailwind	Pitch Angle

Flight analysis 2 – reference speed set incorrectly



P	C	Je	D	Jo
Altitude	Slope-possible	Slope-possible	Slope-possible	Slope-possible
Ground Spd	Descent Rate	Ground Spd	Descent Rate	Altitude Skip
Descent Rate	Tailwind	Descent Rate	Tailwind	Descent Rate
Tailwind	Ground Spd	Tailwind	Ground Spd	Ground Spd
Altitude Skip	Altitude	Altitude	Altitude	Tailwind
Slope-possible	Altitude Skip	Altitude Skip	Altitude Skip	Altitude

Flight analysis 1 – Nominal Flight



Outline

- Background
  - Precursor discovery problem, uses, challenges
- Methodology
  - ADOPT algorithm
- Case Studies
  - Take-off Stall Hazard
  - STAR procedure adherence
- Summary

ADOPT's features

- Data mining based precursor discovery algorithm

- Input
  - Feed in time series data with adverse event
  - Feed in nominal time series data
  - Data could be continuous, categorical, text, images

- Output
  - Precursor time instants
  - Precursor variables
  - Probability score

- Correlation and not Causation

ADOPT's features

- Use any/all domain knowledge
  - Selecting variables
  - Scoping problems in space, time
  - Hand-engineering features
- Use any classifier of choice
  - SVM, decision tree, K-NN, logistic regression
- Extends to multiple adverse events
  - Holistic analysis, safety margins
- Parallelizable
  - Multiple CPUs
  - Analyze multiple airports, airspaces, aircrafts in parallel

Summary

- Precursor discovery is an important problem with uses in multiple applications in Aviation.
- ADOPT is an efficient data mining solution to find precursors.
- Two case studies are presented to show the setup, working and features of ADOPT.
- ADOPT will be open-sourced in the near future.

Thank You